

## Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated:

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.

- If you encounter an accession number from an older search run against UniProt (results file extension .upr) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2006, 14:24:44 ; Search time 234 Seconds

Perfect score: 2316 ; Sequence: 1 AVSTSPAAFPYDQILAAHAA.....NPPLTAALACCIGARA 438

Title: US-10-788-746-6  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%; Maximum Match 100%; Listing first 45 summaries

Database: UniProt 05.80.\*  
1: uniprot\_sprot;\*  
2: uniprot\_trembl;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	2316	100.0	1284	09WJ35_9VIRU
2	2316	100.0	1320	09WJ35_9VIRU
3	2234	96.5	1319	09WJ35_9VIRU
4	2225	96.1	1319	09WJ35_9VIRU
5	2224	96.0	1319	09WJ35_9VIRU
6	2203	95.1	1322	09WJ35_9VIRU
7	2203	95.1	1324	09WJ35_9VIRU
8	859	37.1	167	09WJ35_9VIRU
9	501	21.6	1258	09WJ35_9VIRU
10	499	21.5	1244	09WJ35_9VIRU
11	487	21.0	1248	09WJ35_9VIRU
12	471	20.6	1241	09WJ35_9VIRU
13	473	20.4	1241	09WJ35_9VIRU
14	471	20.3	1248	09WJ35_9VIRU
15	470	20.3	1248	09WJ35_9VIRU
16	470	20.3	1248	09WJ35_9VIRU
17	468	20.2	1248	09WJ35_9VIRU
18	465.5	20.1	1253	09WJ35_9VIRU
19	465	20.1	1247	09WJ35_9VIRU
20	462.5	20.0	1253	09WJ35_9VIRU
21	458.5	19.8	1236	09WJ35_9VIRU
22	456.5	19.7	1241	09WJ35_9VIRU
23	454.5	19.6	1240	09WJ35_9VIRU
24	452.5	19.5	1239	09WJ35_9VIRU
25	452.5	19.5	1239	09WJ35_9VIRU
26	452.5	19.5	1242	09WJ35_9VIRU
27	452.5	19.5	1242	09WJ35_9VIRU
28	452.5	19.5	1242	09WJ35_9VIRU
29	452.5	19.5	1242	09WJ35_9VIRU
30	452.5	19.5	1242	09WJ35_9VIRU
31	452.5	19.5	1242	09WJ35_9VIRU

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%; Maximum Match 100%; Listing first 45 summaries

Database: UniProt 05.80.\*  
1: uniprot\_sprot;\*  
2: uniprot\_trembl;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### ALIGNMENTS

RESULT 1	09WJ35_9VIRU	PRN; 1284 AA.
ID	09WJ35_9VIRU PRELIMINARY;	
AC	09WJ35;	
DT	01-Nov-1999 (Tremblrel. 12, Created)	
DT	01-Nov-1999 (Tremblrel. 12, Last sequence update)	
DT	01-Mar-2004 (Tremblrel. 26, Last annotation update)	
DB	Polyprotein.	
OS	Salmon pancreas disease virus.	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;	
OC	Alphavirus.	
OX	NCBI_TaxID=84589;	
RN	[1] _TaxID=84589;	
RP	NUCLEOTIDE SEQUENCE.	
RX	MEDLINE-9920813; PubMed=10191183; DOI=10.1006/viro.1999.9654;	
RA	Weston J.H., Welsh M.D., McLoughlin M.F., Todd D.	
RT	"Salmon Pancreas Disease virus," an alphavirus infecting farmed Atlantic Salmon ( <i>Salmo salar</i> ).";	
RL	Virology 256:188-195 (1999).	
DR	GO; GO:0016020; C:membrane; I:EA.	
DR	HSSP; P03115; I:VCP.	
DR	EMBL; AU012631; CAB42823.1; -; Genomic_RNA.	
DR	P03115; I:VCP.	
DR	GO; GO:0008233; F:peptidase activity; I:EA.	
DR	GO; GO:0004252; F:serine-type endopeptidase activity; I:EA.	
DR	GO; GO:0005198; F:structural molecule activity; I:EA.	
DR	GO; GO:0006508; P:proteolysis and peptidolysis; I:EA.	
DR	InterPro; IPR02548; Alpha_E1_glycop.	
DR	InterPro; IPR00936; Alpha_E2_glycop.	
DR	InterPro; IPR02533; Alpha_E3_glycop.	
DR	InterPro; IPR00930; Peptidase_S3.	
DR	InterPro; IPR01998; Viral_glycoB_cen.	
DR	Pfam; PF00944; Alpha_core; 1.	
DR	Pfam; PF01589; Alpha_E1_glycop; 1.	
DR	Pfam; PF00943; Alpha_E2_glycop; 1.	
DR	Pfam; PF01563; Alpha_E3_glycop; 1.	
DR	PRINTS; PR00788; TOGA VIRIN.	
SK	Capris Protein Hydrolase; Polyprotein; Protease; Structural protein.	
SEQUENCE	1284 AA; 138915 MW; 496621EB019BRC0 CREG64;	

Query	Match	100.0%; Score 2316; DB 2; Length 1284; Best Local Similarity 100.0%; Pred. No. 1.8e-17; Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 AVSTSPAAFPYDQILAAHAA.....NPPLTAALACCIGARA	61 QSGVTAKGAGAEGSIRYLGDRGKHAADNTRVTTAKCDVQIQTGHYLNGPGQS
Db	P08979_QEBV	414 QSGVTAKGAGAEGSIRYLGDRGKHAADNTRVTTAKCDVQIQTGHYLNGPGQS

QY	2 Q88798_BEEV	121 LTVAATLGDTRHQCTVPEHQVTEKTRERSKGHHLSDWKKCUTPFRPSKALVLDV 180
QY	3 Q88798_BEEV	Q88798 eastern equ

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model  
Run on: April 19, 2006, 14:31:58 ; Search time 48 Seconds  
(without alignments)

Title: US-10-788-746-6  
Perfect score: 2316  
Sequence: 1 AVSTSPRAFYDFTOILAHAA.....NPPPLTALTAALCCCTPGARA 438

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060  
Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%\*  
Listing First 45 summaries

Database : Issued\_Patents\_AAI\*  
1: /cgn2\_6/prodata/1/1aa/5\_COMB\_pep:/\*  
2: /cgn2\_6/prodata/1/1aa/6\_COMB\_pep:/\*  
3: /cgn2\_6/prodata/1/1aa/H\_COMB\_pep:/\*  
4: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB\_pep:/\*  
5: /cgn2\_6/prodata/1/1aa/RB\_COMB\_pep:/\*  
6: /cgn2\_6/prodata/1/1aa/backfilesl.pep:/\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No. Score Query Match Length DB ID Description

Result No.	Score	Query Match	Length	DB ID	Description
1	2316	100.0	438	2 US-09-674-866A-6	Sequence 6, Appli
2	2316	100.0	1359	2 US-09-674-866A-6	Sequence 3, Appli
3	437.5	18.9	1236	2 US-10-023-659A-4	Sequence 4, Appli
4	437.5	18.9	1236	2 US-10-023-49A-6	Sequence 6, Appli
5	402.5	17.4	1253	1 US-07-920-81C-3	Sequence 3, Appli
6	402.5	17.4	1253	2 US-08-466-277-3	Sequence 3, Appli
7	402.5	17.4	1253	2 US-09-688-942-3	Sequence 6, Appli
8	395.5	17.1	1245	1 US-08-801-653A-6	Sequence 6, Appli
9	395.5	17.1	1245	2 US-09-102-048-6	Sequence 6, Appli
10	395.5	17.1	1245	2 US-09-367-64-6	Sequence 6, Appli
11	394	17.0	981	2 US-09-991-558-13	Sequence 6, Appli
12	389.5	16.8	1245	1 US-08-801-363A-3	Sequence 3, Appli
13	389.5	16.8	1245	2 US-09-102-48-3	Sequence 3, Appli
14	389.5	16.8	1245	2 US-09-991-764-3	Sequence 3, Appli
15	383.5	16.6	1245	1 US-08-001-363A-10	Sequence 10, Appli
16	383.5	16.6	1245	2 US-09-102-48-10	Sequence 10, Appli
17	383.5	16.6	1245	2 US-09-367-764-10	Sequence 10, Appli
18	113	4.9	2315	2 US-09-543-661A-5334	Sequence 5434, Appli
19	107	4.6	437	2 US-09-830-189C-2	Sequence 2, Appli
20	107	4.6	437	2 US-09-991-181-355	Sequence 355, Appli
21	107	4.6	437	2 US-09-990-44-355	Sequence 355, Appli
22	107	4.6	437	2 US-10-033-301-16	Sequence 16, Appli
23	107	4.6	437	2 US-10-997-333-355	Sequence 355, Appli
24	107	4.6	437	2 US-09-992-598-355	Sequence 2, Appli
25	107	4.6	437	2 US-09-073-559-2	Sequence 26762, Appli
26	103.5	4.5	298	2 US-09-248-76A-28762	Sequence 2916, Appli
27	99.5	4.3	685	2 US-10-104-047-2916	Sequence 2916, Appli

ALIGNMENTS

RESULT 1  
US-09-674-866A-6  
; Sequence 6, Application US/09674866A  
; Patent No. 6719980el  
; GENERAL INFORMATION:  
; APPLICANT: Akzo NO. 6719980el NV  
; TITLE OF INVENTION: Structural Proteins of Fish Pancreatic Disease Virus  
; TITLE OF INVENTION: and Uses Thereof  
; FILE REFERENCE: 1/98376US  
; CURRENT APPLICATION NUMBER: US/09/674, 866A  
; CURRENT FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: B98201461.5  
; PRIOR FILING DATE: 1998-05-08  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO: 6  
; LENGTH: 438  
; TYPE: PRT  
; ORGANISM: Salmon Pancreatic disease virus  
; FEATURE:  
; OTHER INFORMATION: E2

US-09-674-866A-6  
; Query Match 100.0%; Score 2316; DB 2; Length 438;  
; Best Local Similarity 100.0%; Prod. No. 5.9e-242;  
; Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
; Qy 1 AVSTSPRAFYDFTOILAHAA.....NPPPLTALTAALCCCTPGARA 438  
; Db 1 AVSTSPRAFYDFTOILAHAA.....NPPPLTALTAALCCCTPGARA 438  
; 1. QSGVTAKGAGGETSIRYLGDRGKWAADNTRLYVTTAKCDVQLQATGHYILANGPVGS 60  
; 61 QSGVTAKGAGGETSIRYLGDRGKWAADNTRLYVTTAKCDVQLQATGHYILANGPVGS 120  
; 61 QSGVTAKGAGGETSIRYLGDRGKWAADNTRLYVTTAKCDVQLQATGHYILANGPVGS 120  
; 121 LTVAAATLDGTRHQCTTNPQVTSKTRSKGHLSIDMTKKCCTRSTPKKSALYLVD 180  
; 121 LTVAAATLDGTRHQCTTNPQVTSKTRSKGHLSIDMTKKCCTRSTPKKSALYLVD 180  
; 181 YDALPISVLSITVWTCSDSOCTVVPGTVKFDKCKSADSATVTPTSQDTCCEPV 240  
; 181 YDALPISVLSITVWTCSDSOCTVVPGTVKFDKCKSADSATVTPTSQDTCCEPV 240  
; 181 YDALPISVLSITVWTCSDSOCTVVPGTVKFDKCKSADSATVTPTSQDTCCEPV 240  
; 241 LTAASITQKPHLSAMPGSGKKEVARIIPPFPEPTACRVSAPLPSITYERSDVLIA 300  
; 241 LTAASITQKPHLSAMPGSGKKEVARIIPPFPEPTACRVSAPLPSITYERSDVLIA 300  
; 301 GTAKKPVLLTRNLGPHSNATSEMTQKLRIPPTQGIELTWNANPMPHFWSSYRAS 360  
; 301 GTAKKPVLLTRNLGPHSNATSEMTQKLRIPPTQGIELTWNANPMPHFWSSYRAS 360



Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on:

April 19, 2006, 14:33:40 ; Search time 27 Seconds

173.825 Million cell updates/sec

Title: US-10-788-746-6

Perfect score: 2316

Sequence: 1 AVSTSPRAYFDQIILAHAA.....NPPPLTALTAUCCITPGARA 438

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs; 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%\*

Maximum Match 100%\*

Listing First 45 summaries

Database : Published Applications AA\_New:\*

1: /SIBS/ptodata/1/pubpaal/us08\_NEW\_PUB.pep:\*

2: /SIBS/ptodata/1/pubpaal/us06\_NEW\_PUB.pep:\*

3: /SIBS/ptodata/1/pubpaal/us07\_NEW\_PUB.pep:\*

4: /SIBS/ptodata/1/pubpaal/fct\_NEW\_PUB.pep:\*

5: /SIBS/ptodata/1/pubpaal/us09\_NEW\_PUB.pep:\*

6: /SIBS/ptodata/1/pubpaal/us10\_NEW\_PUB.pep:\*

7: /SIBS/ptodata/1/pubpaal/us11\_NEW\_PUB.pep:\*

8: /SIBS/ptodata/1/pubpaal/us60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Query Match Length DB ID Description

1 107.5 4.6 5291 7 US-11-052-554A-281 Sequence 281, APP

2 107 4.6 437 6 US-10-131-826A-466 Sequence 466, APP

3 107 4.6 437 6 US-10-973-115B-466 Sequence 466, APP

4 107 4.6 437 6 US-10-213-535-16 Sequence 16, APP

5 107 4.6 437 6 US-10-218-784-162 Sequence 162, APP

6 107 4.6 437 6 US-10-219-661-162 Sequence 162, APP

7 107 4.6 437 6 US-10-219-662-162 Sequence 162, APP

8 107 4.6 437 6 US-10-219-664-162 Sequence 162, APP

9 107 4.6 437 6 US-10-233-134-162 Sequence 162, APP

10 107 4.6 437 7 US-11-230-153-466 Sequence 466, APP

11 99.5 4.3 685 7 US-11-072-512-2916 Sequence 2916, APP

12 99.5 4.3 487 6 US-10-745-586-11 Sequence 11, APP

13 97 4.2 403 6 US-10-453-372-1170 Sequence 1170, APP

14 96 4.1 381 7 US-11-087-099-3485 Sequence 3485, APP

15 96 4.1 940 7 US-11-045-004-39 Sequence 39, APP

16 95 4.1 382 6 US-10-793-798B-40 Sequence 40, APP

17 95 4.1 1709 6 US-10-935-561-973 Sequence 973, APP

18 95 4.1 1709 6 US-10-453-372-410 Sequence 410, APP

19 94.5 4.1 334 7 US-11-056-568A-1273 Sequence 11273, APP

20 94.5 4.1 365 7 US-11-056-568A-11272 Sequence 11272, APP

21 94.5 4.1 419 7 US-11-056-568A-11271 Sequence 11271, APP

22 94.5 4.1 583 7 US-11-080-991-64 Sequence 64, APP

23 94 4.1 626 6 US-11-045-004-25 Sequence 25, APP

24 94 4.1 1700 6 US-10-453-372-398 Sequence 398, APP

25 94 4.1 1700 6 US-10-453-372-412 Sequence 412, APP

ALIGNMENTS

RESULT 1

US-11-052-554A-281

; Sequence 281, Application US/11052554A

; Publication No. US20050288866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE PROTEINS

; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 3085/40359A

; CURRENT APPLICATION NUMBER: US/11/052-554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEB/2004

; SOFTWARE: Patentin version 3.3

; SEQ ID NO: 281

; LENGTH: 5291

; TYPE: PRT

; ORGANISM: Escherichia coli 0157:H7

US-11-052-554A-281

Query Match 4.6% ; Score 107.5; DB 7; Length 5291; Best Local Similarity 19.8%; Pred. No. 1.7; Matches 104; Conservative 68; Mismatches 175; Indels 177; Gaps 22;

QY Db 1902 VSAWPAGSVTIAASGSTSAGNPVSPVTHPVTVDSLAVAVSINITADDVINAKBGAALTL 1961

QY Db 1962 SGSTSCVBAQGTVTTFGGKTSATVANGSWSTSPADMAALRGGDASQASVSNNG 2021

QY Db 2022 NSATTHAYSDASAPTVTNTIAGDILNAABAGALITIGS--STAEGOTIVV-T 2076

QY Db 2077 LNTGNTGTGQDGSWSVSADSLTASNYTNAVSDRAGNPASVHMLT----- 2130

QY Db 164 TRFSTPKSALYLVVDYDALPISFISTV---VTCSDSCTVRVPPGTVKDFKKS 219

QY Db 2131 -----VDTSPV-VINTVAGDDVIVNATEHAQAOQIISGAT-----GA 2167

Copyright (C) 1993 - 2006 Biocceleration Ltd. GenCore version 5.1.7

PS	PA	(PURD ) PURD RES FOUND.	XX
XX	XX	McCrory PB, Sanders DA, Jeffers SA, Davidson BL, Sims PL;	XX
CC	XX	WPT, 2005-384668/39.	XX
CC	DR	N-PSDB; ADZ5899.	XX
XX	XX	Introducing a selected nucleic acid sequence into an airway epithelial cell by contacting the airway epithelial cell with a pseudotyped retrovirus, where the pseudotyped retrovirus comprises a glycoprotein and retroviral capsid comprising the nucleic acid sequence.	XX
XX	XX	The invention relates to introducing a selected nucleic acid sequence into an airway epithelial cell. The method comprises contacting the airway epithelial cell with a pseudotyped retrovirus, where the pseudotyped retrovirus comprises a glycoprotein in which a portion of an	XX
DISCLOSURE	SEQ ID NO 8; 102pp; English.	RESULTS	ALIGMENTS
RESULT 1			
AD275898	ID	AD275898 standard; protein; 487 AA.	
XX	AC	ACD275898;	
XX	DT	28-JUL-2005 (first entry)	
XX	DE	Robe river virus E3-E2 protein, seq id 8.	
XX	KW	Gene therapy; retrovirus based gene therapy; cystic fibrosis; airway epithelial cell; E3-E2.	
XX	OS	ROBB River virus.	
XX	PN	US2005112098-Al.	
XX	PD	26-MAY-2005.	
XX	PR	26-MAR-2004; 2004US-00811353.	
XX	PR	26-OCT-2001; 2001US-0353221P.	
PR	PR	26-OCT-2001; 2001US-0356436P.	
PR	PR	04-JUN-2002; 2002US-0346064P.	
PR	PR	28-OCT-2002; 2002US-045834545.	
PR	PR	27-MAR-2003; 2003US-0458070P.	
PR	PR	04-JUN-2003; 2003WO US017577.	
XX	PA	(IOWA ) UNIV IOWA RES FOUND.	XX
XX	PI	McCrory PB, Sanders DA, Jeffers SA, Davidson BL, Sims PL;	XX
XX	DR	N-PSDB; ADZ5899.	XX
XX	XX	Introducing a selected nucleic acid sequence into an airway epithelial cell by contacting the airway epithelial cell with a pseudotyped retrovirus, where the pseudotyped retrovirus comprises a glycoprotein and retroviral capsid comprising the nucleic acid sequence.	XX
XX	XX	The invention relates to introducing a selected nucleic acid sequence into an airway epithelial cell. The method comprises contacting the airway epithelial cell with a pseudotyped retrovirus, where the pseudotyped retrovirus comprises a glycoprotein in which a portion of an	XX

OM protein - protein search, using sw model					
Run on: April 19, 2006, 14:28:04 ; Search time 42 Seconds (without alignments) 1003.403 Million cell updates/sec					
Title: US-10-788-746-6					
Perfect score: 2316					
Sequence: RAVSTSPAFYDFTQILAAHAA.....NPPIPTALTAACCGPGRRA 438					
Scoring table: BLOSUM62					
Searched: Gapop 10.0 , Gapext 0.5					
283416 seqs, 9216763 residues					
Total number of hits satisfying chosen parameters: 283416					
Minimum DB seq length: 0					
Maximum DB seq length: 200000000					
Post-processing: Minimum Match 0% Maximum Match 100% Listing first 45 summaries					
Database : PIR 80:*					
1: Pirl:*					
2: pir2:*					
3: pir3:*					
4: pir4:*					
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.					
SUMMARIES					
Result No.	Score	Query Match Length	DB ID.	Description	
1	436.5	19.7	141	2 S26373	genome polypeptide
2	454.5	19.6	1240	1 VHWVTEV	structural polypeptide
3	452.5	19.5	1242	2 S72350	structural polypeptide
4	452	19.5	1247	1 VHWVN2	structural polypeptide
5	445.5	19.3	422	1 VHWV48	structural polypeptide
6	442.5	19.1	1242	2 A5605	structural polypeptide
7	437.5	18.9	1239	1 VHWTE	structural polypeptide
8	434.5	18.8	1236	1 VHWVTE	structural polypeptide
9	416.5	18.0	1254	1 VHWVRA	structural polypeptide
10	413.5	17.9	422	1 VHWV70	structural polypeptide
11	399	17.2	1254	1 JQ1978	structural polypeptide
12	398.5	17.2	1253	1 VHWV	structural polypeptide
13	397.5	17.2	1245	1 VHWV82	structural polypeptide
14	396	17.1	1254	1 JQ1979	structural polypeptide
15	396	17.1	1255	1 D44213	structural polypeptide
16	395	17.1	1254	1 VHWVTT	structural polypeptide
17	394.5	17.0	755	2 S42462	structural polypeptide
18	393.5	17.0	1145	2 S37136	structural polypeptide
19	393	17.0	1254	1 VHWVTE	structural polypeptide
20	386.5	16.7	1245	1 VHWVB	structural polypeptide
21	386.5	16.7	1245	1 VHWVB2	structural polypeptide
22	384	16.6	1255	1 D44213	structural polypeptide
23	383.5	16.6	423	1 VHWVSB	structural polypeptide
24	126	5.4	69	2 G37264	E2 glycoprotein -
25	115	5.0	69	2 F37264	E2 glycoprotein -
26	109.5	4.7	472	2 T05852	hypothetical protein
27	109	4.7	1777	2 T34569	hypothetical protein
28	107.5	4.6	5188	2 B85547	probable RNR family protein
29	107.5	4.6	5291	2 F90696	hypothetical protein

## ALIGNMENTS

ALIGNMENTS	
RESULT 1	
S26373	genome polyprotein - eastern equine encephalomyelitis virus
NRContains: 6K protein; capsid protein B1; envelope protein B2; envelope protein	protein C; envelope protein
C.Species: eastern equine encephalomyelitis virus	
C.Date: 06-Jan-1994 #sequence_revision 01-Nov-1996 #text_change 31-Dec-2004	
C.Accession: S26373	
R; Volkchova, V.B.; Volkchova, V.A.; Notesov, S.V.	
Mol. Gen. Mikrobiol. Virusol. 5, 8-15, 1991	
A.Title: Complete nucleotide sequence of the eastern equine encephalomyelitis virus genome	
A.Reference number: S26369; PMID:91375524; PMID:1896061	
A; Accession: S26373	
A; Molecule type: mRNA	
A; Residues: 1-1241 <VOL>	
A; Cross-references: UNIPROT:Q66579; UNIPARC:UP1000001639; EMBL:X63135; NID:959185; PR	
C; Note: sequence could not be checked because of bad print in paper	
C; Superfamily: pestivirus genome polyprotein	
C; Keywords: capsid protein; envelope protein; glycoprotein; polyprotein	
F; 1-260/Product: capsid protein C #status predicted <CAP>	
F; 261-323/Product: envelope protein B3 #status predicted <EP3>	
F; 244-433/Product: envelope protein B2 #status predicted <EP2>	
F; 744-800/Product: 6K protein #status predicted <6K>	
F; 801-1241/Product: envelope protein B1 #status predicted <EP1>	
Query Match	19.7%; Score 456.5; DB 2; Length 1241;
Best Local Similarity	30.2%; Pred. No. 5.3e-27;
Matches	133; Conservative 64; Mismatches 189; Indels 55; Gaps 17;
Qy	21 ASPRAYKPCDCGACISRIAIIDWVSSSDHILRMRVQESQSGTAKGGAGGETSRYL- 79
Db	335 ARPYTADCGCNGHSHRCDSPIAIBEVRGDAHAGVIRIOTSAMFLGKTDG----VOLAIMS 389
Qy	80 ---GRDGKTHAADMTRLWRRTAKCDVQLOAHTGHYLANCEPGVQGOSTIATVAATLDGTHQCTT 136
Db	390 PAKGKTKOKSIIKDN-LAVRTSACPSLVSHGHTVILQOCPGDTWTFGHDGPNTCTV 447
Qy	137 VPERQW----TEKFRTRSKGHHLSDMTKKCTRFRSTPDKKSAKLVDYDALLPSVLI 190
Db	448 A--HKVEFRPVGRKSYRHPPEHGVTL---PCNRY--THKRDQHRYVEMHQ-Q-PGLVAD 497
Qy	191 STVTCSDQGTCVTRPGTIVKFKCKSADSATVPTFSDQSOTFCBEPFLTAASITQK 250
Db	498 HSLSIHSIAKVKUTPSAQKVYKCKPDPREGT---TSDYTTCTDVQCRAYLIDNK 554
Qy	251 PHL-RSAMPLPSG-GKEVKARIKPFPPPEPATCRVSVAPLSDITYTRESDVILAGTAKYKPV 308
Db	555 KWVINSGRPLRGEGTDTPKGKLUHVPPVVKACIAALAPFLVERHGRTILHLYDAPL 614
Qy	309 LTTNLGKPSNATEWQIQLQKTRLPVPTQGIEUTWGNVAPMHRFSSVRYASGDDAYP 368
Db	615 LTTNLGKPSNATEWQIQLQKTRLPVPTQGIEUTWGNVAPMHRFSSVRYASGDDAYP 673